VizBBA - Visualization of Bike Sharing Rides in Bay Area

Charan Teja
chg81@pitt.edu

Hui Zhang
huz26@pitt.edu

Indraneil Bardhan
inb11@pitt.edu

Shuning Tong
sht72@pitt.edu

ABSTRACT

BikeShare systems have been rapidly growing in popularity across the world. This project aims to explore the Bay Area BikeShare data to analyze the patterns and trends of bike usage using effective visualization techniques. VizBBA tries to answer questions including popular stations, trip flow distribution, peak days and hours, relationship between weather and bike usage, and subscriber behavior. The system will help users gain insights on the rebalancing problem and help policy makers decide where to build more bike stations.

Keywords

Information Visualization, BikeShare.

1. INTRODUCTION

Many successful BikeShare programs have measured decrease in trips made by automobiles and reduced traffic conditions. They have become an indispensable part of modern culture.

Bay Area BikeShare system consists of a fleet of specially designed, heavy-duty, very durable bikes that are locked into a network of docking stations located throughout the region. Bay Area bikes can be rented from and returned to any station in the system, creating an efficient network with many possible combinations of start and end points. [1] With hundreds of bikes at stations, the system is available for use 24 hours a day, 365 days a year. Bikes stations are built in San Francisco, San Jose, Palo Alto and Mountain View.

We have chosen to work with Bay Area BikeShare data because the program in collaboration with Ford Motor Company is expanding tenfold from 700 to 7,000 bikes starting in Spring 2017. In the pursuit of finding new station locations, the BikeShare program has outreached to the public to find their opinion by conducting various workshops and feedback programs in which it has received an overwhelming response of 5000 suggestions for stations.

In this paper, we have examined the official open data [2] using visualization techniques to address some questions including the relationship between weather and bike usage, popular stations, trip flow distribution, peak days and hours, and subscriber behavior. Our project will not only help individuals and groups who are working on BikeShare research professionally, but also provide a clear demostration of BikeShare patterns to everyday users who are interested in BikeShare stories. Figure 1 is a teaser of VizBBA system. As shown, it aims to answer the question of popular stations and trip flow distribution. More questions will be answered in other visualizations shown later.

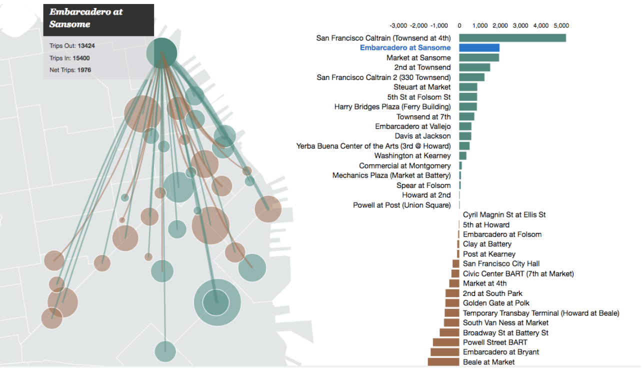


Figure 1 Teaser visualization of VizBBA system. A flow map with graduated symbol is shown on the left side to present trip flow distribution, and a bar chart is shown on the right side to present popular incoming and outgoing stations. These two visualizations are connected. When users mouse hover a circle on flow map, the corresponding bar will be highlighted. Vice versa.

2. DATASET

2.1 Why We Choose

The dataset examined in this project is extracted from the Bay Area BikeShare website [2]. The data is openly available for public use. While looking for datasets, we have also looked carefully at bike sharing data from Pittsburgh, Chicago, New York and Washington DC. Looking at the data, we realized that data from the San Francisco Bay Area already contained the weather data in its dataset. Having worked on the BikeShare projects before, we knew that getting weather data and moulding it into a required shape would be a time-consuming task. While choosing the data, we believed that this would allow us to concentrate on making visualizations.

2.2 What It Looks Like

We have three years of data. We decided to focus on the most recent year: September 2015 - August 2016. It mainly consists of four parts:

2.2.1 Trip Data

The trip data includes trip id, duration, start date, start station, start terminal, end date, end station, end terminal, bike number, subscription type and user's zip code.

2.2.2 Station Data

The station data includes station id, station name, latitude, longitude, dock count and installation time.

2.2.3 *Status Data*

The status data includes station id, available bike number, available dock number, time.

2.2.4 *Weather Data*

The weather data includes date, temperature, humidity, visibility, wind speed, cloud cover, events, wind direction and many other weather indicators.

After examination, we decided to use trip, station and weather data. The status data is way too large since it includes status for every minute. It's definitely very important, but we figured that it's not related with what we try to answer.

Also, we decided to use only San Francisco data since the majority of stations (60%) and trips (92%) are registered in San Francisco.

3. RELATED WORK

Various works have been created using data from Bay Area BikeShare Program to visualize bike usage patterns. In this section, we will introduce some of the related works that caught our eyes and inspired us to design work of our own. Limitations of each design will also be explained briefly to showcase possible adjustments that can be done to make our visualization stand out from previous works. The data analyzed in related work came from the same dataset, Bay Area BikeShare open data from August 2013 to February 2014.

3.1 Adjacency Matrix

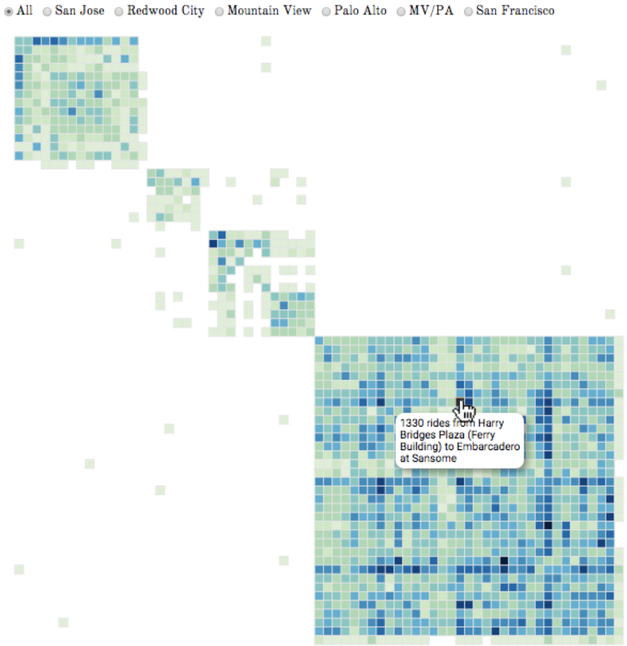


Figure 2 Adjacency matrix generated using D3.js showing amount of trips recorded by starting station and ending station. An apparent drawback is that it does not show the station labels. It's difficult for a normal user to understand the meaning of this matrix.

One possible solution showcased by Tyler Field [3] is the use of adjacency matrix (See Figure 2). It uses color to depict the data, darker hue representing larger amount of trips.

This work appears to be a good representation for bike trips since human visual system favors color pre-attentively. This work also manages to group those bike stations in the same city together, enabling easy-to-detect cluster effect. We can see from this visualization that most of the bike trips have starting station and ending station in the same city. Only a few of them are cross cities (those light green small rectangles dissociated from the clusters). A second finding is that San Francisco has the largest amount of bike trips, as the largest rectangle shown in the graph. This finding supports our decision of focusing on San Francisco only. What's more, this work provides mouse hover interaction. When users mouse hover one small rectangle, a detailed information will be shown including the number of rides, from station and end station.

We realized several drawbacks of this work except missing labels. The main issue is that it's not easy to find a geographic pattern of bike trips. It only shows the name of stations and the number of rides. Even if I'm a frequent San Francisco BikeShare user, there is a high chance that I'm not familiar with the names of all stations. From this visualization, users get to know there are 1330 rides from Harry Bridges Plaza to Embarcadero at Sansome in that period. But where are these two stations? Another drawback of this work is that the order of stations in the matrix is not meaningful. We do not find a reasonable explanation of why stations are ordered this way. For matrix representations, order of data entries impacts the appearance a lot. That's why we order books by their class for political books dataset. Here, the order of stations looks random, thus the pattern shown here is also random.

3.2 Interactive Map

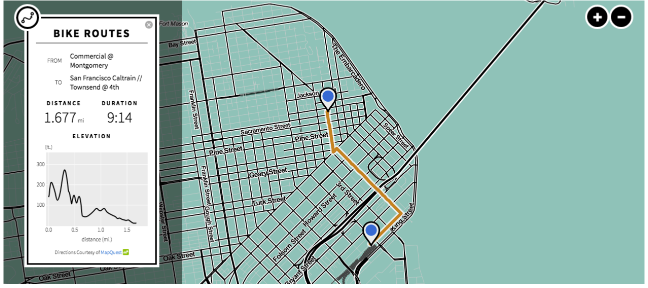


Figure 3 Interactive map generated using Leaflet and D3.js. It brings together different pieces of information that are of interest to its users, including trip stats, availability, bike routes and nearby photos. Users can choose from these four modes and see the detailed information creator organized in the side panel. This figure only shows the bike route mode.

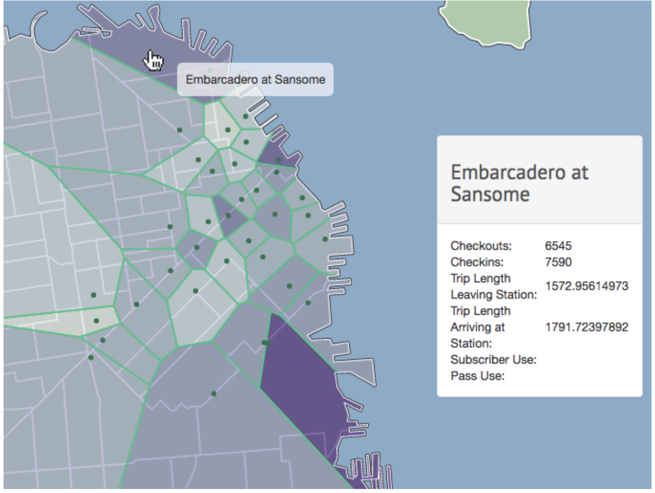
As shown in Figure 3, this work is an interactive map showing detailed information when users choose from four modes: trip stats, availability, bike routes, and nearby photos. This work is from Virot Ta Chiraphadhanakul [4]. It is in the form of node-link representation since icons represents stations and lines represents routes.

Compared with the adjacency matrix, this work allows users to see the locations of stations. Another notable advantage of this work is that it coordinates all the detailed information in the side panel.

While we were inspired by the node-link representation and the select interaction offered in this work, we also noticed some drawbacks. The visualization is good for finding information for one specific station or route, but might not be so effective in giving users a more general look. Users may want to compare different stations, especially for policy makers who are dealing with the rebalancing problem. But this work does not provide a visual encoding for users to compare stations directly.

Another drawback of this work is that the detailed information provided in the side panel is in the form of either lineplot or barplot. The number of visual encodings is too few compared with the information it carries. This work is great for professionals, but might not make everyday users feel inspired.

3.3 Voronoi Diagram



Embarcadero at Sansome

Checkouts: 6545
Checkins: 7590
Trip Length Leaving Station: 1572.95614973
Trip Length Arriving at Station: 1791.72397892
Subscriber Use:
Pass Use:

Figure 4 Voronoi diagram generated using D3.js showing the division of regions of San Francisco based on the BikeShare station the space is closest to. Each cell represents the total amount of bike check-outs and check-ins. Darker cells represent stations that have more activity.

David Belford and Jennifer Wong [5] created voronoi diagram to show bike usage in different area, calculated based on the station the area is closest to. Points are used to mark the location of stations. Polygons are used to mark different areas. Color is used to encode the amount of bike check-outs and check-ins. Users can mouse hover a polygon or a point to see the name of that station, and when users click a polygon or a point, a detailed information panel will be shown.

One advantage of this work is that it uses very clear visual encodings to tell the story the creators want to convey. The visual encodings and interactions used in this work will allow normal users understand very well. Also, it provides a good overview which makes comparison easier, and give detailed information on demand.

The creators mentioned that the stations near Caltrain and along the piers get the most usage. They thought it suggesting some demand among commuters and tourists. We found out that this finding does not require a voronoi diagram. The amount of bike usage can be encoded on the points using size. A possible reason the creators use voronoi diagram here is that they assume people who check in or check out from one station is living nearby. But that may not be the case. In San Francisco, people can take BART to commute a long distance to downtown, and then take bikes to reach their final destinations. So the activity in one station does not represent the exact activity situation of cyclists living nearby. It may contain activity of users living far away.

This visualization inspires our interests in answering cyclists related questions, which will be illustrated later.

4. VISUAL DESIGN

4.1 Design Process

A good visualization can not be achieved without reasonable design and arrangement ahead of time. In this section, we will talk about how the design of our project developed throughout the process.

We applied the guideline of visual storytelling for this project. Kosara and Mackinlay [6] mentioned, stories can help effectively present information and make a point in a memorable way. Visualizing the stories of bike usage will make our work more logically structured.

As mentioned in the previous section, the main target audiences for our project consists of professional policy makers and normal people with limited knowledge.

Since the target audiences have been determined, the next step is to consider what to communicate to them using our visualizations. We put ourselves in users' shoes and figured out some questions they want to know about bike usage.

VizBBA system contains four parts, each part answering one question. They are basically Where, When, Weather and User (cyclist).

1. Where are the popular stations for incoming and outgoing?

2. When do trips usually occur in one week and one day?

3. Does the weather impacts bike usage?

4. Where are those neighborhoods with high cyclist activity?

The design process of each step can be found below.

4.1.1 Where

Given the rides dataset, one of the questions we wanted to address was the popularity of the BikeShare stations. We want to understand which bike stations are the most popular and most heavily used in the system. Since one bike ride has a start station and an end station, we decided to find out the most popular stations for outgoing and incoming.

Since we aim to present the geospatial information, a map will be a natural choice. Inspired by Virot Ta Chiraphadhanakul's interative map, we plan to use a node-link representation. A flow map came to our mind. Nodes will be BikeShare stations, edges will be trips. This visualization will help us analyze popular stations and popular routes. As mentioned in related work part, we want to add more visual encodings to the map to make it more friendly to normal users. So we decided to add graduated symbol on the flow map to show the size of net incoming trips or net outgoing trips. To help users quickly find the ranking of all the stations, we decided to add a barplot. The flow map with graduated symbols, combined with barplot, will answer the question in a clear manner.

4.1.2 When

Since we planed to explore the weekly and hourly patterns of bike trips, we quickly found out that this is a time series visualization and the arrangement of time is cyclic. So we decided to use circular heat chart with color encoding the amount of bike trips 7 days per cycle to detect periodic weekly pattern and 24 hours per cycle to detect periodic hourly pattern.

4.1.3 Weather

It's commonly known that the weather in San Francisco is very unique, due to its closeness to the Pacific Ocean. We want to find out if unique weather in San Francisco affects how people commute using bikes. Visibility (due to fog), wind speed and temperature become the three factors we aim to explore.

This is also time series visualization, and the arrangement is linear. We decided to use lineplot for this question, since lines are the best to encode correlations.

4.1.4 User

What is the geographical distribution of BikeShare subscribers? We started asking this question. But soon, we found out that there is no information regarding individual subscribers in our dataset. All we have is trip and the zipcode of the subscriber who made that trip. Since individual subscribers can take multiple rides, it is not feasible to accurately know the number of subscribers. So we change the question a little bit. Where are those neighborhoods (divided by zipcode) with high cyclist (subscriber) activity?

There is another user group named "customer" in our dataset. This group of users are only registered for a short term BikeShare pass (1 to 3 days). We do not use their data because their place of residence often locate outside of Bay Area. In data processing, we also drop all those data with zipcode outside Bay Area.

Naturally, a map is a perfect choice for visualizing this kind of dataset with geographical information. Inspired by David Belford and Jennifer Wong's voronoi diagram, we thought that a choropleth map is a good solution for our project. After an evaluation of the preliminary design, we came to realize that we need to add another symbol map layer on top of it to show the stations. The number of subscriber trips from specific zip code area will be shown using colors. And bike stations will be plotted on the map using points.

By visualizing the neighborhoods where subscribers live and the stations, we aim to explore the relationship between the neighborhoods of residence and the neighborhoods where the bike stations locate. We aim to show how the number of rides taken by subscribers differs in neighborhoods with bike stations in contrast to neighborhoods without bike stations. We are expecting to find some area has few bike stations and less BikeShare subscribers, thus indicating the space of developing more subscribers when more stations are built there.

4.2 Visual Encodings

4.2.1 Where

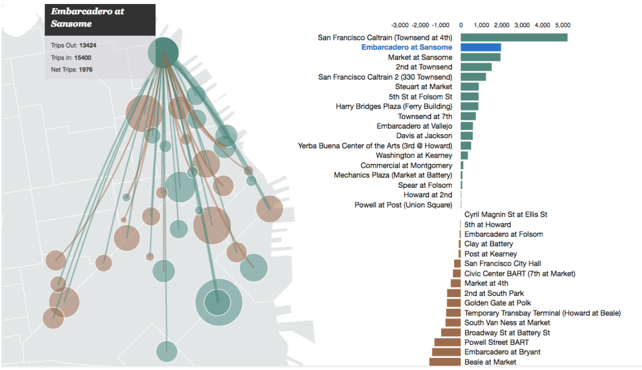


Figure 5 Flow map with graduated symbols combined with bar chart. Green color represents the stations with more incoming trips. Yellow color represents the stations with more outgoing trips. The size of circles, the length of bars, the weight of lines all represent the amount of net incoming or outgoing bike trips.

As mentioned in class, bars are the best encodings for ranking relationship. The length of the bars encodes the net incoming trips or net outgoing trips: abs(incoming trips - outgoing trips). Users will be able to find the most popular stations for incoming (in this case, San Francisco Caltrain) and outgoing (in this case, 2nd at Folsom). And it will be easy for them to know the order of popular stations simply from the bar chart.

Position is the second visual encoding here to show the geospatial information of all the bike stations.

Size of the circles is the third visual encoding here to show basically the same meaning of the length of bars. But duplicates does not mean useless. It allows users to compare different stations directly on geo map.

Line weight is the fourth visual encoding here to show the net incoming or net outgoing trips for each route.

Last but not least, color encodes whether there is more incoming trips (green) or outgoing trips (yellow). It represents the status of each station, which is critical to solve the rebalancing problem. That's why we use color (the most pre-attentive visual encoding) to present this information.

We put ourselves into the shoes of policy makers. Now, simply by looking at the flow map and the bar chart, we can determine which station is over checked out and which station is over checked in. What's more, we know their locations. So we can perform rebalancing operation based on these information. For example, we notice Embarcadero at Sansome is over checked in, and its surrounding stations are mostly over checked out. So as a policy maker, we can make rebalancing policy that moves some of the bikes from Embarcadero at Sansome to its surrounding stations periodically.

4.2.2 When

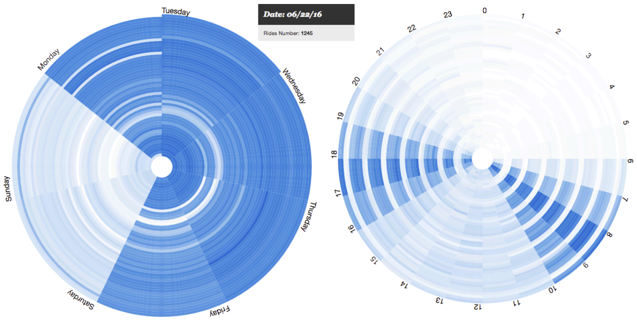


Figure 6 Circular heat charts showing the weekly and hourly patterns of bike trips. The left circle has a whole year's data. The right circle has data of 50 days from September 2015. Darker blue stands for higher amount of bike trips.

Saturation is used here to encode the amount of bike trips of one day (left circle) or one hour (right circle). We put ourselves into the shoes of normal users. We can easily detect that there is more darker blues in the left circle from Monday to Friday and there is more darker blues in the right circle from 7 to 10 in the morning and from 4 to 7 in the evening, indicating higher bike usage.

This time series visualization has some properties that worth discussing. The scope of time is interval-based. The arrangement of time is cyclic. Each small segment represents the state of bike trips in that time period. And it's univariate. Only one variable (amount of bike trips) is shown here along with time.

4.2.3 Weather

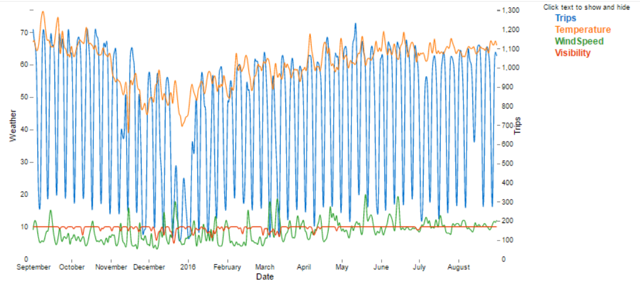


Figure 7 Line plots showing the relationship between weather and bike usage. Blue line represents the amount of trips. Yellow line stands for temperature. Green line indicates wind speed. Red line showcases visibility.

Lines are the best encoding for correlations. Here the blue line represents the dependent variable we want to look at. Other three lines stand for three independent variables. Color is used here to encode variable type. Trip and weather conditions are encoded using data positions. We put ourselves into the shoes of normal users. We can easily detect that where temperature drops, trip amount also drops. The same with wind speed and visibility.

This time series visualization also has some properties that worth discussing. The scope of time is also interval-based since the granularity is per day. The arrangement of time is linear. Each data point stands for the state of bike trips or weather in that time period. And it's multivariate. Four variables are shown here along with time.

4.2.4 User

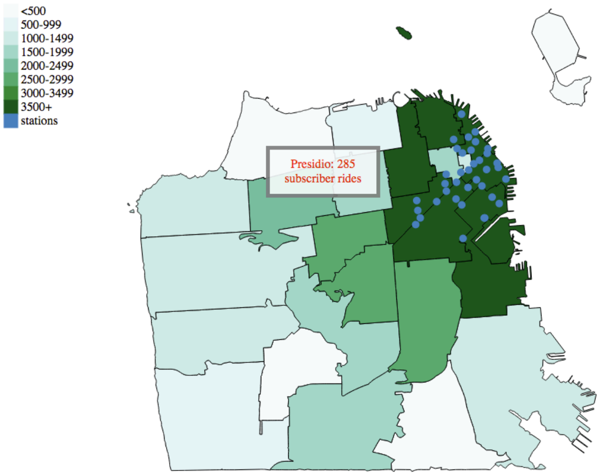


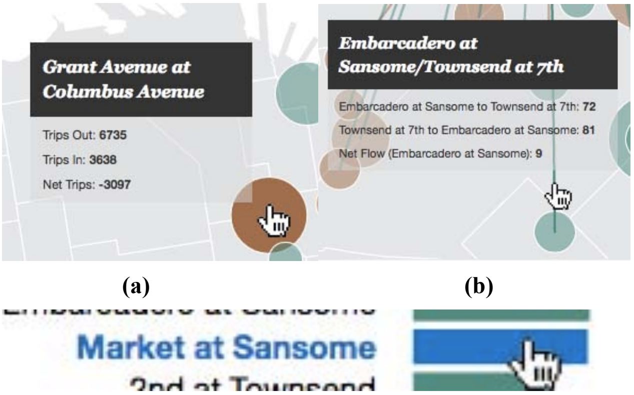
Figure 8 Choropleth map representing neighborhoods with subscriber (cyclist) activity. Darker green represents higher cyclist activity. Blue dots represents the bike stations.

Saturation is used here to encode the number of rides made by subscribers live in one area. Since the distribution of data is heavily skewed, so we adjust the scale to let more area be marked dark green. The highest amount of bike rides for one area appears to be about 28,000 and the lowest amount is about 300. The scale is set manually to let those zip code areas with more than 3500 bike rides per year be marked dark green. Position of points is used here to encode the geospatial information of bike stations.

We put ourselves into the shoes of policy makers. We can easily detect that those neighborhoods that are close to the bike stations tend to have high cyclist activity, indicating the space of developing more subscribers (an untapped population) when more stations are built in low activity neighborhoods.

4.3 Interactions

4.3.1 Where


Grant Avenue at Columbus Avenue

Trips Out: 6735

Trips In: 3638

Net Trips: -3097

Embarcadero at Sansome/Townsend at 7th

Embarcadero at Sansome to Townsend at 7th: 72

Townsend at 7th to Embarcadero at Sansome: 81

Net Flow (Embarcadero at Sansome): 9

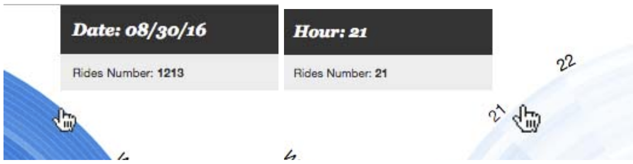
(a) (b)

(c)

Figure 9 Flow map with graduated symbols combined with bar chart. (a) Mouse hover one circle to see detailed information of this station. Corresponding bar will be highlighted. (b) Mouse click one circle to see a bunch of lines be drawn representing trip flows. Mouse hover one line to see detailed information of this flow. (c) Mouse hover one bar to highlight it. Corresponding circle will be highlighted and detailed information will be shown.

Interactions included in this visualization are: select (click one circle), highlight (mouse hover one circle or bar), coordinate (flow map and bar chart are connected). In information centric view, it meets users' intent to select (mark something as interesting), explore (show other details), abstract (show less details) and connect (show related items).

4.3.2 When


Date: 08/30/16

Rides Number: 1213

Hour: 21

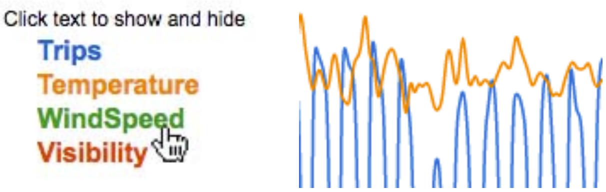
Rides Number: 21

(a) (b)

Figure 10 Circular heat charts. (a) Mouse hover one segment of the weekly circle to see the date and number of rides in that date. (b) Mouse hover one segment of the hourly circle to see the hour and the number of rides in that hour.

The only interaction included in two circular heat charts is highlight (mouse hover segments to show details). In information centric view, it meets users' intent to explore and abstract.

4.3.3 Weather

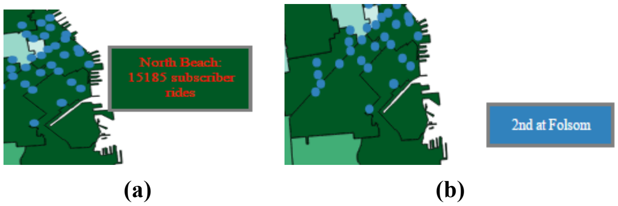


(a) (b)

Figure 11 Line plots. (a) Mouse click text to show or hide the current line. (b) Lines will be filtered according to user's selection.

Interactions included in the line plots are: filter and select (mouse click one text to show or hide current line). In information centric view, it meets users' intent to select, explore and filter (show something conditionally).

4.3.4 User



(a) (b)

Figure 12 Choropleth map. (a) Mouse hover one polygon to see the name and the rides made by subscribers live in that neighborhood. (b) Mouse hover one point to see the name of the station.

The only interaction included in the choropleth map is highlight (mouse hover polygon or point to see details). In information centric view, it meets users' intent to select, explore and abstract.

5. IMPLEMENTATION NOTES

We use D3.js only to generate our visualizations. Since we have very clear questions organized for our visualizations, we use simple annotated charts to be the structure of our system. Large headlines and brief introductions are used to be the layout of our VizBBA system. Buttons are used to navigate from one question to another.

There were many challenges in the implementation process. One of the main challenges we faced while generating the choropleth map is that it is very difficult to obtain maps that are demarcated by zip code areas. The zip code areas are decided by the US Postal Service and keep changing. For example, some zip codes that have been put out of use by the USPS are still used by a small number of people. We currently use San Francisco Zip Codes data [7] from the government website. Combining the layer of neighborhoods with the layer of stations is another challenge, which we managed to solve with the help of tutorials [8] on the internet.

For line plots part, we tried to build a model and draw these four lines at one time but failed since their scales are different. Then we changed the method by drawing lines one by one. For interactions, we tried to add mouse hover function to show all four lines data on same day at the same time. Unfortunately, we didn’t find enough references to implement this function. We can only show one line’s data at a time, so we omit this function and it will be our future work.

6. EVALUATION

We perform an evaluation regarding our VizBBA visualization system. Questions include:

1. Overall, how effective is the system? (scale 1 to 5)
2. Does the system convey the information that you are looking for? (scale 1 to 5)
3. Does the system inspire your thinking? (scale 1 to 5)
4. How do you rate the ease of interaction with the system? (scale 1 to 5)
5. Are the colors used in the system friendly and easy to differentiate? (scale 1 to 5)
6. How do you rate the ease of navigation within the system? (scale 1 to 5)
7. What do you like most about this visualization?
8. What suggestions do you have?

We got feedback from 12 users. The average credit is 4.62 for effectiveness, 4.50 for information convey, 4.73 for inspiration, 4.14 for the ease of usage, 4.14 for colors and 4.45 for navigation. We receive 3 votes for circular heat charts as their favorite, 2 votes for both interaction and use of colors, 1 vote for flow map and 1 vote for choropleth map. One suggestion is that the line plots is difficult to follow because of the huge variation of data points. It would be better if we can plot aggregate weekly data instead of daily data to avoid jumping weekends.

7. DISCUSSION

The biggest advantage of our VizBBA system is that it answers the questions we put forth in the beginning. We are motivated by the expanding plan of Bay Area BikeShare program. And we aimed at solving the rebalancing problem. These two problems are well solved by our system. The second advantage is that it uses multiple visual encodings to allow better user experience. As mentioned in related work part, our work is based on previous visualizations and we make improvement by providing general picture first and then show details on demand.

The disadvantage of our system is mentioned in the evaluation part.

In the future, we hope to look at this question from the point of view of casual customers and explore how they use the bike sharing system.

8. CONCLUSION

We have made a great effort to create the VizBBA system to visualize the pattern of bike usage in San Francisco. We use a flow map with graduated symbols combined with bar chart, circular heat charts, line plots, and choropleth map. Two of them are time series visualizations. Two of them are geovisualizations. The system solves the rebalancing problem and answers the four questions users may seek answers for. The system will help normal users gain insights on bike sharing related problems and help policy makers decide where to build more bike stations.

9. ACKNOWLEDGMENTS

Our thanks to ACM SIGCHI for allowing us to modify templates they had developed. Our thanks to Bay Area BikeShare Program for allowing us to use their data and utilize some of the visualizations from 2014 Open Data Challenge. We also use the zip code data from San Francisco government. The open source D3 design developed by Peter Cook also helped us develop our own circular heat charts. We rely a lot on Mike Bostock and Scott Murray's D3 tutorials. Finally, we want to thank Dr. Yuru Lin for offering the wonderful course.

10. CONTRIBUTIONS

10.1 Shuning Tong

Shuning is the creator of the flow map with graduated symbols combined with bar chart and the circular heat charts. She also is the organizer of the presentation and paper.

10.2 Hui Zhang

Hui is the creator of line plots. He also is the organizer of the web layout of VizBBA system. He contributes to preparing presentation and paper.

10.3 Indraneil Bardhan

Neil is the creator of choropleth map. He also contributes to preparing presentation and paper.

10.4 Charan Teja

Charan is the creator of force-directed graph (not used in the final system). He also contributes to preparing presentation and paper.

11. REFERENCES

[1] About Bay Area Bike Share. http://www.bayareabikeshare.com/about Accessed 2016-12-13.

[2] Open Data. http://www.bayareabikeshare.com/open-data Accessed 2016-12-13.

[3] Tyler F. 2014. Bay Area BikeShare Open Data Challenge. *Best Analysis*. http://thfield.github.io/babs/ Accessed 2016-12-13.

[4] Virot T. C. 2014. Bay Area BikeShare Open Data Challenge. *Best Overall Visualization*. http://babs.virot.me/ Accessed 2016-12-13.

[5] David B., and Jennifer W. 2014. Bay Area BikeShare Open Data Challenge. http://www.wongjennifer.com/maps/bike-station-map/ Accessed 2016-12-13.

[6] Kosara, R., and Mackinlay, J. 2013. Storytelling: The next step for visualization. Computer, 46(5), 44-50. doi:10.1109/MC.2013.36

[7] San Francisco Zip Codes. https://data.sfgov.org/Geographic-Locations-and-Boundaries/San-Francisco-ZIP-Codes/srq6-hmpi Accessed 2016-12-13.

[8] Mike B. 2012. Let's Make a Map. https://bost.ocks.org/mike/map/ Accessed 2016-12-13.

[9] Mike B. 2013. *How Selections Work.*https://bost.ocks.org/mike/selection/

[10] Marlin F. *Horizontal Stacked Bar Chart for D3.js*. https://jsfiddle.net/datashaman/rBfy5/4/light/

[11] Mike B. 2012. *Object Constancy*. https://bost.ocks.org/mike/constancy/

[12] Leon du T. 2015. Unemployment Ranked with Horizontal Bars. http://bl.ocks.org/leondutoit/6436923

[13] Ian J. 2015. Simple Transitions with D3.js. http://bl.ocks.org/enjalot/1429426

[14] Mike B. 2016. Multi-Series Line Chart. https://bl.ocks.org/mbostock/3884955

[15] Bay Area BikeShare Open Data Challenge. 2014.http://doi.acm.org/10.1145/161468.16147http://www.bayareabikeshare.com/open-data

[16] Color Brewer 2.0: 2015.http://colorbrewer2.org/. Accessed 2016-12-13.

[17] Scott M. D3 Tutorials. http://alignedleft.com/tutorials/d3/ Accessed 2016-12-13.